

What is Claimed is:

1. A computerized method of predicting future behavior of an individual, the method comprising:
using a computer program to analyze the content of internet websites already visited by that individual.
2. A method according to Claim 1, further comprising combining text from a plurality of the visited websites, identifying a plurality of the most informative words of that text, and using data representative of those most informative words as inputs to an automated predictive model whose outputs indicate the individual's likely future behavior.
3. A method according to Claim 2, further comprising identifying, for words of the combined text, their frequency of occurrence in the combined text and also of their occurrence in a large text corpora in the same language, and selecting as the said most informative words those whose said frequency of occurrence is significantly greater in the combined text than in the large text corpora.
4. A method according to Claim 3, comprising identifying, from a database of semantic vectors derived from co-occurrence statistics, the semantic vector of each of the said most informative words, and using the semantic vectors as the said representative data.
5. A method according to Claim 4, wherein the number of the most informative words is a predetermined number appropriate to give sufficient predictive accuracy in a reasonable amount of computation time.

6. A method according to Claim 5, further comprising varying the said predetermined number of most informative words in order to determine its optimum, by refitting the predictive model for each value of the number and noting the predictive accuracy and the time taken.

7. A method according to Claim 6, further comprising determining the predictive accuracy by a cross-validation procedure.

8. A computerized method carried out by a business in relation to its customers or potential customers as individuals for customer relationship management, the method comprising:

analyzing the content of internet websites already visiting by customers;
predicting the customers' future behavior including their commercial requirements relating to that behavior; and
then communicating appropriately with selected ones of those customers.

9. A computer program for predicting future behavior of an individual, the program comprising:

means for analyzing the content of internet websites already visited by that individual.

10. A computer program for customer relationship management, the program comprising:

means for analyzing the content of internet websites already visited by customers; and

means for predicting those customers' future behaviors including their commercial requirements relating to those behaviors.

11. A computer program according to Claim 10, further comprising means for allowing a business operating the program to communicate appropriately with selected ones of those customers.

12. A computer program according to Claim 11, further comprising means for combining text from a plurality of the visited internet websites, to identify a plurality of the most informative words of that text, and to use data representative of those most informative words as inputs to an automated predictive model whose outputs indicate the individual's likely future behavior.

13. A computer program according to Claim 12, further comprising means for identifying, for words of the combined text, their frequency of occurrence in the combined text and also of their occurrence in a large text corpora in the same language, and means for selecting as the said most informative words those whose said frequency of occurrence is significantly greater in the combined text than in the large text corpora.

14. A computer program according to Claim 13, further comprising means for identifying, from a database of semantic vectors derived from co-occurrence statistics, the semantic vector of each of the said most informative words, and using the semantic vectors as the said representative data.

15. A computer system for executing the computer program of Claim 9.

16. A computer program for customer relationship management carried out by a business in relation to its customers or potential customers as individuals for customer relationship management, the computer program comprising:

means for analyzing the content of internet websites already visiting by customers;

mean for predicting the customers' future behavior including their commercial requirements relating to that behavior; and

means for communicating appropriately with selected ones of those customers.

17. A computer system for executing the computer program of Claim 16.